

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
20 January 2005 (20.01.2005)

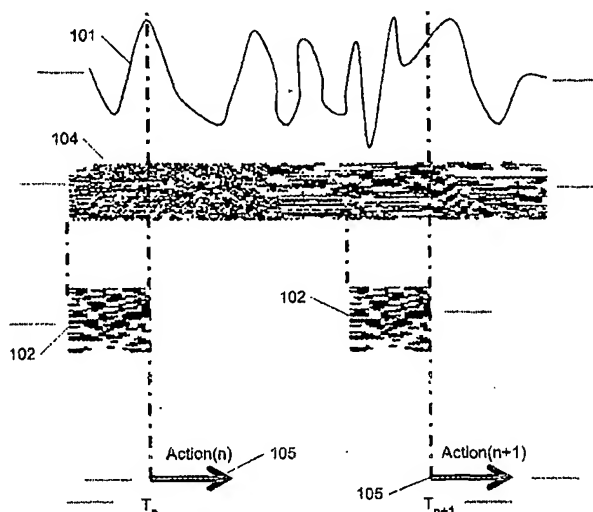
PCT

(10) International Publication Number
WO 2005/006758 A1

- (51) International Patent Classification⁷: **H04N 7/24**,
G06T 1/00, H04N 7/26 [GB/GB]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (21) International Application Number: PCT/IB2004/051128 (74) Agent: GROENENDAAL, Antonius, W., M.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).
- (22) International Filing Date: 5 July 2004 (05.07.2004) (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: 03102116.5 11 July 2003 (11.07.2003) EP
- (71) Applicant (for all designated States except US): KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): OOSTVEEN, Job, C. [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). TEN KATE, Warner, R., T. [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). DENISEN, Adrianus, J., M. [NL/NL]; c/o Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL). ROBERTS, David, K. (84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: METHOD AND DEVICE FOR GENERATING AND DETECTING A FINGERPRINT FUNCTIONING AS A TRIGGER MARKER IN A MULTIMEDIA SIGNAL



(57) Abstract: This invention relates to a device and a method of relating one or more trigger actions with a multimedia signal and corresponding method and device for detecting one or more trigger actions in a multimedia signal. One fingerprint is generated on the basis of a segment of the multimedia signal at each trigger time point. The generated fingerprint(s) is stored in a database and communicated to playback devices. During playback fingerprints of the multimedia signal are generated and matched against fingerprints in the database. When a match is found, associated trigger actions is retrieved. The trigger time point is determined as a time point near or at a segment of the multimedia signal that caused a fingerprint to match. In this way, a simple and reliable way of detecting trigger markers for trigger actions is enabled without modifying the multimedia signal.

WO 2005/006758 A1

**Declaration under Rule 4.17:**

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE,

IT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

Published:

- with international search report
— before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

Method and device for generating and detecting a fingerprint functioning as a trigger marker in a multimedia signal

The invention relates to a method, and a corresponding device, of detecting trigger instants/information in a multimedia signal. The invention also relates to a method, and a corresponding device, of associating trigger information with a multimedia signal. Further, the invention relates to a computer readable medium having stored thereon instructions for causing one or more processing units to execute the method according to the invention.

10

A current trend is to enhance passive television viewing and/or music listening on a given playback device by creating more interactive programs and/or listening experiences or by "connecting" external actions to a piece of video and/or audio content. As one simple example, a commercial can be enhanced by embedding a URL to a web site with further information, where the URL can be extracted and retrieved by the playback device. In order to facilitate such a function it is necessary to enable a reliable detection of time points in a television program, a movie, a music piece, etc. where such additional information is relevant.

Examples of situations where such additional information is useful or interesting in connection with a broadcast program are:

- trigger&link: (see e.g. W. ten Kate et.al. "trigger&link – A new dimension in television program making", Lecture Notes in computer Science, vol. 1242, pp51-65, Springer, 1997) trigger&link allows interactivity in television programs. In addition to the normal program, additional information concerning specific segments of the program is available to the viewer through a different distribution channel. At the start of a given segment that is associated with an enhancement (additional information) an icon is displayed, alerting the viewer that additional information may be viewed on his TV. For instance, at the appearance of an actor in a movie, some biographic data of the actor may be made available. In the user terminal (e.g. a set top box, etc.) the icons

are overlayed on the video at the relevant time instants, thereby requiring these instants to have been indicated in the video stream.

- Local Insertion: During a national broadcast, specific parts of the program may be replaced by a regional program in some regions. For instance, some advertisements may be replaced by advertisements for local shops, or, in a news show some regions may have their local weather forecast rather than the national one. The national program producer can indicate which segments are suitable for such local insertion. At the local redistribution site (e.g. at the cable head-end), the indicated segments may be replaced by local content.

In both of the above situations, it is necessary to mark or associate specific time instants in the video stream at which additional information should be available. At these time instants the receiver should be triggered to perform or provide some kind of action. This may be done by such mechanisms as DSM-CC in MPEG/DVB. However, this requires the broadcaster's cooperation to insert these triggers thereby making an enhancement service provider dependent on the broadcaster.

One previously known way of performing time marking in a video stream is e.g. using fields of the MPEG transport stream structure that can be used to hold the marking information.

Another previously known way is using a blanking interval. In analog distribution, the marking information can be embedded in the vertical blanking interval or in the inactive video lines.

Both of the above known ways need the cooperation of all actors in the broadcast chain to make sure that the marking information is not destroyed before the signal arrives at its destination. For instance, in case of the MPEG solution, a re-multiplexing operation could easily remove information that is written in the user data fields in the stream. Moreover, every decoding and successive re-encoding step would certainly not retain this information. In case of the use of the vertical blanking for carrying the trigger information, the situation is even more difficult, as actors in the broadcast chain might write other information at the same position (the vertical blanking is used for many things and there is no uniform agreement about the control over usage of the blanking interval). Also, standards converters (like PAL-NTSC) and other equipment in the broadcast chain may not retain all information in the vertical blanking interval.

Yet another way is using watermarking. A watermark may be embedded in the video frames at the relevant time instants. The Philips Watercast System is, among others,

being sold for this purpose. A disadvantage of watermarking is the fact that it necessarily changes the video/audio.

5 It is an object of the invention to provide a method and corresponding device of relating one or more trigger actions with a multimedia signal and corresponding method and device for detecting one or more trigger actions in a multimedia signal that solves the above-mentioned problems. A further object is to provide this in a simple and efficient way. Another object is to enable simple, reliable and accurate localisation of a given part of a
10 multimedia signal. A further object is to enable detection of trigger actions without modifying the multimedia signal.

 This is achieved by a method (and corresponding device) of relating one or more trigger actions with a multimedia signal, the method comprising the steps of

- 15 - providing at least one trigger time point and for each trigger time point providing at least one representation of least one associated trigger action, where each trigger time point indicates a time point of the multimedia signal for which the at least one associated trigger action is to be available during playback of the multimedia signal, and
- 20 - for each given trigger time point deriving a fingerprint on the basis of a segment of the multimedia signal, where the segment of the multimedia signal is unambiguously related with the given trigger time point,
- and by a method (and corresponding device) of detecting one or more trigger actions in a multimedia signal, the method comprising the steps of:
 - 25 - generating a fingerprint stream on the basis of the multimedia signal,
 - comparing a segment of the fingerprint stream with one or more fingerprints stored in a second database in order to determine if a match exists or not between the segment and a fingerprint in the second database, the second database further comprising for each stored fingerprint at least one representation of at least one associated action, and
 - 30 - if a match exists retrieving the at least one representation of the at least one associated action associated with the matching fingerprint.

 In this way, a simple and efficient way of handling time markers in a multimedia signal for given actions is obtained. A fingerprint thereby serves as a trigger of a particular action, event, etc. associated with a particular point in time of the multimedia

signal. Further, this is enabled without the multimedia signal needing to be modified. Additionally, the time marking detection is time independent as it is dependent on the specific content of the multimedia signal only thereby avoiding problems if a multimedia signal being a television program or the like is delayed.

5 A fingerprint of a multimedia object/content/signal is a representation of perceptual features of the object/content/signal part in question. Such fingerprints are sometimes also known as "(robust) hashes". More specifically, a fingerprint of a piece of audio or video is an identifier which is computed over that piece of audio or video and which does not change even if the content involved is subsequently transcoded, filtered or otherwise
10 modified.

 Preferably, the derived fingerprint is an audio and/or video fingerprint. Alternatively, animations and/or streaming text, etc. is used as a source for creating a fingerprint.

 Advantageous embodiments of the methods and devices according to the
15 present invention are defined in the sub-claims.

 Further, the invention also relates to a computer readable medium having stored thereon instructions for causing one or more processing units to execute the method according to the present invention.

20

 Figure 1a schematically illustrates generation of fingerprint(s) used as trigger marker(s) according to the present invention.

 Figure 1b schematically illustrates detection and use of fingerprint(s) as trigger marker(s) according to the present invention.

25

 Figure 2 illustrates a schematic block diagram of a fingerprint generation device according to the present invention;

 Figure 3 illustrates a schematic block diagram of a playback device detecting and using fingerprints according to the present invention;

 Figure 4 illustrates one example of tables/records according to the present
30 invention.

 Figure 1a schematically illustrates generation of fingerprint(s) used as trigger marker(s) according to the present invention.

Shown is a digital or analog multimedia signal (101) comprising video and/or audio information/content, where one or more 'trigger' actions (forth denoted actions) is to be associated/related with the multimedia signal (101) at certain given 'trigger' time points (forth denoted time points). The one or more actions associated with each time point is to be available, i.e. triggered, at that given particular time point ($T_n; T_{n+1}$) during playback on a playback device. The notation ' $(T_n; T_{n+1})$ ' for a given time point signifies that the time point may be either the shown time point T_n or the shown time point T_{n+1} or in general any suitable (not shown) time point of the signal (101). The associated actions of multiple time points may be the same, different and/or a mix thereof.

The action(s) to be presented/triggered at a given time point may e.g. comprise retrieving and displaying additional information on a display (e.g. presenting biography data for an actor being shown by the multimedia signal, presenting a selectable URL to a web site containing additional information, etc.), retrieving and playing additional information via a speaker, playing another multimedia signal instead of said multimedia signal (101) for a predetermined or variable period of time (e.g. a local weather forecast, a local commercial, etc.) and/or the like. Other examples of action(s) are e.g. stopping/pausing, e.g. temporarily, display/play, executing other control commands, and/or preparing the system for user input(s), e.g. once the trigger action is executed the system waits (for some time) for a specific action of the user. If the trigger action was not executed, the user input will not have any influence. For example, in interactive games the user may only submit his answer after the trigger action has fired/been executed.

For each time point ($T_n; T_{n+1}$) a fingerprint (102) is generated on the basis of a part, segment, etc. (forth denoted segment) of the multimedia signal (101), where the segment of the multimedia signal (101) is unambiguously related with the given time point ($T_n; T_{n+1}$). Preferably, the segment of the multimedia signal (101) is unambiguously related with the given time point ($T_n; T_{n+1}$) by letting the segment of the multimedia signal (101) ending substantially at the given time point ($T_n; T_{n+1}$). In alternative embodiments, the segment of the multimedia signal (101) may start substantially at the given time point ($T_n; T_{n+1}$), the segment of the multimedia signal (101) is starting or ending at a predetermined distance before or after the given trigger time point ($T_n; T_{n+1}$), or the given time point ($T_n; T_{n+1}$) may be at a predetermined time point between a start and an end of the segment of the multimedia signal (101).

The size of the fingerprints and/or the segments may both be of a predetermined fixed size or alternatively of a variable size.

One method for computing a robust fingerprint is described in European patent application 01200505.4, although of course any method for computing a robust fingerprint can be used.

European patent application 01200505.4 describes a method that generates
5 robust fingerprints for multimedia content such as, for example, audio clips, where the audio clip is divided in successive (preferably overlapping) time intervals. For each time interval, the frequency spectrum is divided in bands. A robust property of each band (e.g. energy) is computed and represented by a respective fingerprint bit.

Multimedia content is thus represented by a fingerprint comprising a
10 concatenation of binary values, one for each time interval. The fingerprint does not need to be computed over the whole multimedia content, but can be computed when a portion of a certain length has been received. There can thus be plural fingerprints for one multimedia content, depending on which portion is used to compute the fingerprint over.

Further, video fingerprinting algorithms are known, e.g. from the following
15 disclosure: Job Oostveen, Ton Kalker, Jaap Haitsma: "Feature Extraction and a Database Strategy for Video Fingerprinting". 117-128. IN: Shi-Kuo Chang, Zhe Chen, Suh-Yin Lee (Eds.): Recent Advances in Visual Information Systems, 5th International Conference, VISUAL 2002 Hsin Chu, Taiwan, March 11-13, 2002, Proceedings. Lecture Notes in Computer Science 2314 Springer 2002.

20 According to the present invention, a fingerprint (102) is generated for each time point on the basis of a given segment of the multimedia signal (101) at or near the specific time point.

In this way, a given fingerprint (102) is a trigger marker enabling a very accurate and very precise location of a given time point of the signal (101) without using the
25 specific time point but instead using (a segment of) the signal. Further, this is enabled without changing the signal. For video fingerprinting the localisation is typically frame accurate, at least as long as any distortion of the video signal is not too severe.

After a fingerprint (102) has been generated it is stored for later use in a database, memory, storage and/or the like.

30 There are several advantages in storing fingerprints (102) for a multimedia signal (101) in a database instead of the multimedia signal itself. To name a few: The memory/storage requirements for the database are reduced.

The comparison of fingerprints is more efficient than the comparison of the multimedia signals themselves, as fingerprints are substantially shorter than the signals.

Searching in a database for a matching fingerprint is more efficient than searching for a complete multimedia signals, since it involves matching shorter items.

Searching for a matching fingerprint is more likely to be successful, as small changes to a multimedia signal (such as encoding in a different format or changing the bit rate) do not affect the fingerprint.

Alternatively, if the multimedia content is in the form of more than a single signal, e.g. a separate audio signal and a separate video signal, the fingerprint(s) (102) may be generated on the basis of a single of them (audio or video) or on both.

The generated fingerprints (102) stored in the database may then be distributed to playback devices via the Internet or in a side-channel of a broadcast channel or via some other channel or other means for use during playback according to the present invention. As other examples of distribution is e.g. physical distribution on a storage medium or in a non-electronic way, e.g. requiring the user to enter the fingerprints and actions manually into the playback device.

In a preferred embodiment, a representation of the associated action(s) is also stored for each fingerprint in the database. These representations are preferably also sent to the playback devices. In an alternative embodiment, representations are not stored in the database or used at all when generating the fingerprints. Another party may then provide the representations to the relevant playback devices as well as a relationship between each fingerprint and its associated action(s).

Figure 1b schematically illustrates detection and use of fingerprint(s) as trigger marker(s) according to the present invention. Shown is a digital or analog multimedia signal (101) comprising video and/or audio information/content, where the signal (101) is played back by a suitable playback device. Further shown is a fingerprint stream (104) that is generated continuously or substantially continuously on the basis of the multimedia signal (101). Alternatively, the fingerprint stream (104) is generated in segments. The fingerprint stream (104) (or segments) is compared with fingerprints (102) stored in a database. The stored fingerprints (102) generated as explained in connection with Figure 1a at a production site. The database preferably also comprises representations of the one or more associated actions (105) for each stored fingerprint (102). The stored fingerprints (102) are e.g. received via the Internet or in a side-channel of a broadcast channel or via some other channel or other means from the distribution site. The representations of the associated action(s) (105) may also be received like this. In an alternative embodiment, the representations as well as a

relationship between each fingerprint and its associated action(s) (105) are provided by another party.

When a match between a segment of the fingerprint stream (104) and a given fingerprint (102) in the database is found the representation(s) of the associated action(s) (105) of that particular fingerprint (102) is retrieved and executed at the appropriate time point ($T_n; T_{n+1}$). When a match between a segment of the fingerprint stream (104) and a fingerprint (102) in the database, the appropriate time point ($T_n; T_{n+1}$) is also determined when the fingerprints (102) have been generated as explained in connection with Figure 1a. Preferably, the given time point ($T_n; T_{n+1}$) is determined by letting the segment of the multimedia signal (101) that the matching fingerprint originally has been based on during generation (according to Figure 1a) ending substantially at the given time point ($T_n; T_{n+1}$). In alternative embodiments, the segment of the multimedia signal (101) may start substantially at the given time point ($T_n; T_{n+1}$), the segment of the multimedia signal (101) is starting or ending at a predetermined distance before or after the given trigger time point ($T_n; T_{n+1}$) or the given time point ($T_n; T_{n+1}$) may be at a predetermined time point between a start and an end of the segment of the multimedia signal (101). The playback device simply needs to be aware of the relationship between a given fingerprint and the given time point used during generation.

When a matching fingerprint (102) is determined, the associated one or more actions is also retrieved. The playback device may then execute these actions or present them to a user e.g. awaiting user confirmation before executing them.

The above-mentioned European patent application 01200505.4 describes various matching strategies for matching fingerprints computed for an audio clip with fingerprints stored in a database.

Further European patent application 01202720.7 describes an efficient method of matching a fingerprint representing an unknown information signal with a plurality of fingerprints of identified information signals stored in a database to identify the unknown signal. This method uses reliability information of the extracted fingerprint bits. The fingerprint bits are determined by computing features of an information signal and thresholding said features to obtain the fingerprint bits. If a feature has a value very close to the threshold, a small change in the signal may lead to a fingerprint bit with opposite value. The absolute value of the difference between feature value and threshold is used to mark each fingerprint bit as reliable or unreliable. The reliabilities are subsequently used to improve the actual matching procedure.

A further advantage of the present invention is that if for any reason the broadcast is delayed, the fingerprint matching ensures that the trigger actions still appear at the correct corresponding moment in the broadcast since the invention is time-independent but content-dependent.

5 Figure 2 illustrates a schematic block diagram of a fingerprint generation device according to the present invention. Shown is fingerprint generation device (200) comprising a multi-media signal input module (201), a fingerprinting module (202), a data base, memory storage and/or the like (203) communicating via a bus (205) or the like under the control of one or more microprocessors (not shown). The fingerprint generation device
10 (200) may in one embodiment optionally also comprise a transmitter and receiver (204) for communicating with other systems, devices, etc. via a wired and/or wireless network e.g. like the Internet.

 The multi-media signal input module (201) receives multimedia content e.g. in the form of an analog or digital audio and/or video signal and feeds the multimedia content to
15 the fingerprinting module (202). The fingerprinting module (202) computes a fingerprint on the basis of the received multi-media content. A fingerprint may be derived for the entire content or for a part of the content. Alternatively, several fingerprints may be derived each from a different part. According to the present invention, a fingerprint is derived each time that a trigger action is needed, i.e. for each time point (T_n ; T_{n+1}), as explained in connection
20 with Figure 1a. A representation of the time point(s) is also supplied to the fingerprinting module (202).

 The fingerprinting module (202) then supplies the computed fingerprint(s) to the database (203) preferably together with the associated one or more actions for each fingerprint. As shown in Figure 4, the database (203) comprises fingerprints 'FP1', 'FP2',
25 'FP3', 'FP4', 'FP5', etc. and respective associated actions 'A1', 'A2', 'A3', 'A4', 'A2, A1', etc.

 The database (203) can be organized in various ways to optimize query time and/or data organization. The output of the fingerprinting module (202) should be taken into account when designing the tables in the database (203). In the embodiment shown in Figure
30 4, the database (203) comprises a single table with entries (records) comprising respective fingerprints and associated (sets) of actions.

 Figure 3 illustrates a schematic block diagram of a playback device detecting and using fingerprints according to the present invention. Shown is a playback device (300) comprising a multimedia signal receiver (301), a fingerprint detector (302), a display/play

circuit (303), a data base, memory storage and/or the like (203') communicating via a bus (205) or the like under the control of one or more microprocessors (not shown). The playback device (300) may in one embodiment optionally also comprise a transmitter and receiver (204) for communicating with other systems, devices, etc. via a wired and/or wireless network e.g. like the Internet.

The multimedia signal receiver (301) receives the multimedia signal e.g. in the form of an analog or digital audio and/or video signal to be displayed and/or played e.g. from a broadcasting cable, antenna, satellite dish, etc. arrangement (not shown). The received multimedia signal is fed to the fingerprint detector (302) that derives a fingerprint stream or segments thereof and determines if there are any matches with fingerprints stored in the database as explained in connection with Figure 1b. If a match is found then a representation of the associated action(s) is also retrieved. The appropriate time point for the associated action(s) is given by the matching fingerprint as described above.

The received multimedia signal is displayed and/or played by the a display/play circuit (303) and at the appropriate time point(s) the associated action(s) is executed or presented to a user e.g. awaiting user confirmation before executing the action(s).

Preferably, the data layout of the database (203') corresponds to the one shown in Figure 4.

The playback device (300) may also comprise a buffer mechanism (not shown) for buffering a part of the multimedia signal before displaying/playing it.

Figure 4 illustrates one example of tables/records according to the present invention. Shown is a table comprising fingerprints (102) 'FP1', 'FP2', 'FP3', 'FP4', 'FP5', etc. and respective associated actions (105) 'A1', 'A2', 'A3', 'A4', 'A2, A1', etc. One or more actions (105) are stored for each fingerprint (102). A given fingerprint (102) is only stored in the table once.

In the claims, any reference signs placed between parentheses shall not be constructed as limiting the claim. The word "comprising" does not exclude the presence of elements or steps other than those listed in a claim. The word "a" or "an" preceding an element does not exclude the presence of a plurality of such elements.

The invention can be implemented by means of hardware comprising several distinct elements, and by means of a suitably programmed computer. In the device claim enumerating several means, several of these means can be embodied by one and the same item of hardware. The mere fact that certain measures are recited in mutually different

dependent claims does not indicate that a combination of these measures cannot be used to advantage.

CLAIMS:

1. A method of relating one or more trigger actions with a multimedia signal (101), the method comprising the steps of
 - providing at least one trigger time point (T_n ; T_{n+1}) and for each trigger time point (T_n ; T_{n+1}) providing at least one representation of least one associated trigger action (105),
5 where each trigger time point (T_n ; T_{n+1}) indicates a time point of the multimedia signal (101) for which the at least one associated trigger action is to be available during playback of the multimedia signal (101), and
 - for each given trigger time point (T_n ; T_{n+1}) deriving a fingerprint (102) on the basis of a segment of the multimedia signal (101), where the segment of the multimedia signal
10 (101) is unambiguously related with the given trigger time point (T_n ; T_{n+1}).
2. A method according to claim 1, characterized in that the method further comprises for each obtained trigger time point (T_n ; T_{n+1}), storing the derived fingerprint (102) and the at least one representation of the at least one associated trigger action (105) in a first
15 database (203).
3. A method according to claims 1, 2, characterized in that the one or more derived fingerprints (102) and/or the at least one representation of at least one associated trigger action (105) for the multimedia signal (101) is transmitted to a playback-device (300)
20 via the Internet or in a side-channel of a broadcast channel or via some other channel or means.
4. A method according to claims 1 to 3, characterized in that the segment of the multimedia signal (101) is unambiguously related with the given trigger time point (T_n ; T_{n+1})
25 according to:
 - the segment of the multimedia signal (101) ending substantially at the given trigger time point (T_n ; T_{n+1}),
 - the segment of the multimedia signal (101) starting substantially at the given trigger time point (T_n ; T_{n+1}),

- the segment of the multimedia signal (101) starting or ending at a predetermined distance before or after the given trigger time point (T_n ; T_{n+1}), or
- the given trigger time point (T_n ; T_{n+1}) being at a predetermined time point between a start and an end of the segment of the multimedia signal (101).

5

5. A method of detecting one or more trigger actions in a multimedia signal (101), the method comprising the steps of:

- generating a fingerprint stream (104) on the basis of the multimedia signal (101),
- comparing a segment of the fingerprint stream (104) with one or more fingerprints (102) stored in a second database (203') in order to determine if a match exists or not between the segment and a fingerprint (102) in the second database (203'), the second database (203') further comprising for each stored fingerprint (102) at least one representation of at least one associated action (105), and
- if a match exists retrieving the at least one representation of the at least one associated action (105) associated with the matching fingerprint (102).

15

6. A method according to claim 5, characterized in that said method further comprises the step of: executing the at least one associated action (105) associated with the matching fingerprint (102) at an appropriate trigger time point (T_n ; T_{n+1}).

20

7. A method according to claim 6, characterized in that the appropriate trigger time point (T_n ; T_{n+1}) is given by an unambiguously relation with a segment of a multimedia signal (101) used during generation of the matching fingerprint (102).

25 8. A method according to claims 1 to 4 or claims 5 to 7, characterized in that said multimedia signal (101) is an audio signal, a video signal or a combined audio/video signal.

9. A method according to claims 1 to 4 or claims 5 to 8, characterized in that said at least one associated trigger action (105) is selected from the group of:

- retrieving and displaying additional information on a display,
- retrieving and playing additional information via a speaker,
- playing another multimedia signal instead of said multimedia signal (101) for a predetermined or variable period of time,
- stopping/pausing, e.g. temporarily, display/play,

30

- executing other control commands, and/or
- preparing the system for user inputs.

10. A method according to claims 1 to 4 or claims 5 to 9, characterized in that the
5 derived fingerprint (102) and/or the fingerprint (102) in the second database (203') is an
audio and/or video fingerprint (102).

11. A multimedia device (200) for relating one or more trigger actions with a
multimedia signal (101), the device comprising
10 - means (202; 204) for providing at least one trigger time point (T_n ; T_{n+1}) and for each
trigger time point (T_n ; T_{n+1}) providing at least one representation of at least one
associated trigger action (105), where each trigger time point (T_n ; T_{n+1}) indicates a
time point of the multimedia signal (101) for which the at least one associated trigger
action is to be available during playback of the multimedia signal (101), and
15 - a fingerprint generator (202) adapted to for each given trigger time point (T_n ; T_{n+1})
deriving a fingerprint (102) on the basis of a segment of the multimedia signal (101),
where the segment of the multimedia signal (101) is unambiguously related with the
given trigger time point (T_n ; T_{n+1}).

20 12. A device according to claim 11, characterized in that the device further
comprises a first database (203) having stored the derived fingerprint (102) and the at least
one representation of the at least one associated trigger action (105) for each obtained trigger
time point (T_n ; T_{n+1}).

25 13. A device according to claims 11, 12, characterized in that the device further
comprises a transmitter (204) for transmitting the one or more derived fingerprints (102)
and/or the at least one representation of at least one associated trigger action (105) for the
multimedia signal (101) to a playback-device (300) via the Internet or in a side-channel of a
broadcast channel or via some other channel or means.

30

14. A device according to claims 11 to 13, characterized in that the segment of the
multimedia signal (101) is unambiguously related with the given trigger time point (T_n ; T_{n+1})
according to:

- the segment of the multimedia signal (101) ending substantially at the given trigger time point ($T_n; T_{n+1}$),
- the segment of the multimedia signal (101) starting substantially at the given trigger time point ($T_n; T_{n+1}$),
- 5 - the segment of the multimedia signal (101) starting or ending at a predetermined distance before or after the given trigger time point ($T_n; T_{n+1}$), or
- the given trigger time point ($T_n; T_{n+1}$) being at a predetermined time point between a start and an end of the segment of the multimedia signal (101).

10 15. A audio and/or video playback device (300) for detecting one or more trigger actions in a multimedia signal (101) comprising:

- means (302) for generating a fingerprint stream (104) on the basis of the multimedia signal (101),
- means (302) for comparing a segment of the fingerprint stream (104) with one or
- 15 more fingerprints (102) stored in a second database (203') in order to determine if a match exists or not between the segment and a fingerprint (102) in the second database (203'), the second database (203') further comprising for each stored fingerprint (102) at least one representation of at least one associated action (105), and
- 20 - means (302) for, if a match exists, retrieving the at least one representation of the at least one associated action (105) associated with the matching fingerprint (102).

16. A device according to claim 15, characterized in that said device further comprises: means (303) for executing the at least one associated action (105) associated with

25 the matching fingerprint (102) at an appropriate trigger time point ($T_n; T_{n+1}$).

17. A device according to claim 16, characterized in that the appropriate trigger time point ($T_n; T_{n+1}$) is given by an unambiguously relation with a segment of a multimedia signal (101) used during generation of the matching fingerprint (102).

30

18. A device according to claims 11 to 14 or claims 15 to 17, characterized in that said multimedia signal (101) is an audio signal, a video signal or a combined audio/video signal.

19. A device according to claims 11 to 14 or claims 15 to 18, characterized in that said at least one associated trigger action (105) is selected from the group of:

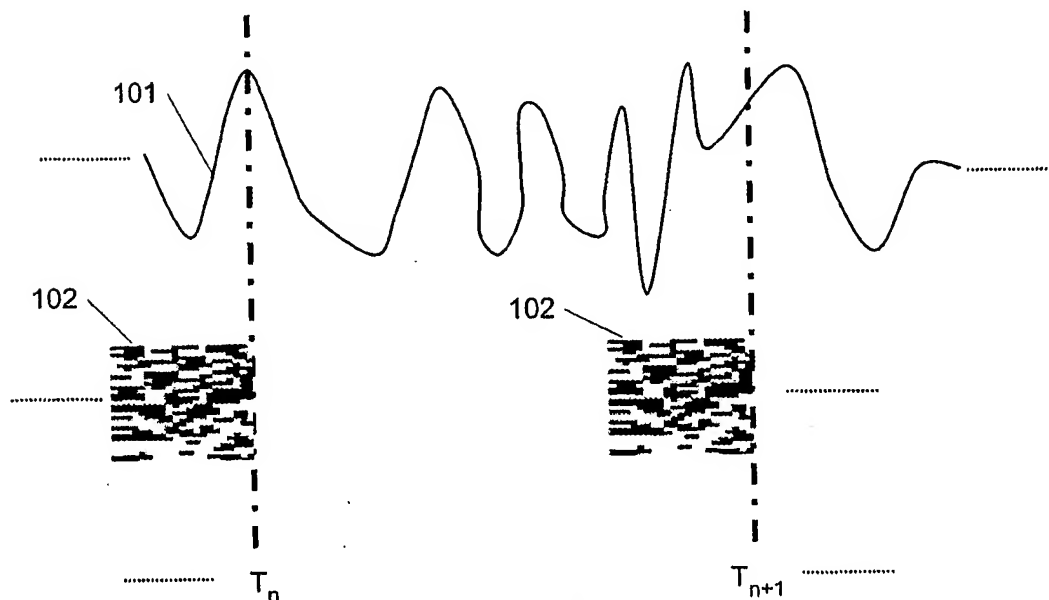
- retrieving and displaying additional information on a display,
- retrieving and playing additional information via a speaker,
- 5 - playing another multimedia signal instead of said multimedia signal (101) for a predetermined or variable period of time,
- stopping/pausing, e.g. temporarily, display/play,
- executing other control commands, and/or
- preparing the system for user inputs.

10

20. A device according to claims 11 to 14 or claims 15 to 19, characterized in that the derived fingerprint (102) and/or the fingerprint (102) in the second database (203') is an audio and/or video fingerprint (102).

15 21. A computer readable medium having stored thereon instructions for causing one or more processing units to execute the method according to any one of claims 1 to 4 or any one of claims 5 to 10.

1/3

**Figure 1a**

| | |
|-----|-------|
| 102 | 105 |
| FP1 | A1 |
| FP2 | A2 |
| FP3 | A3 |
| FP4 | A4 |
| FP5 | A2,A1 |
| ... | ... |

203/203'**Figure 4**

2/3

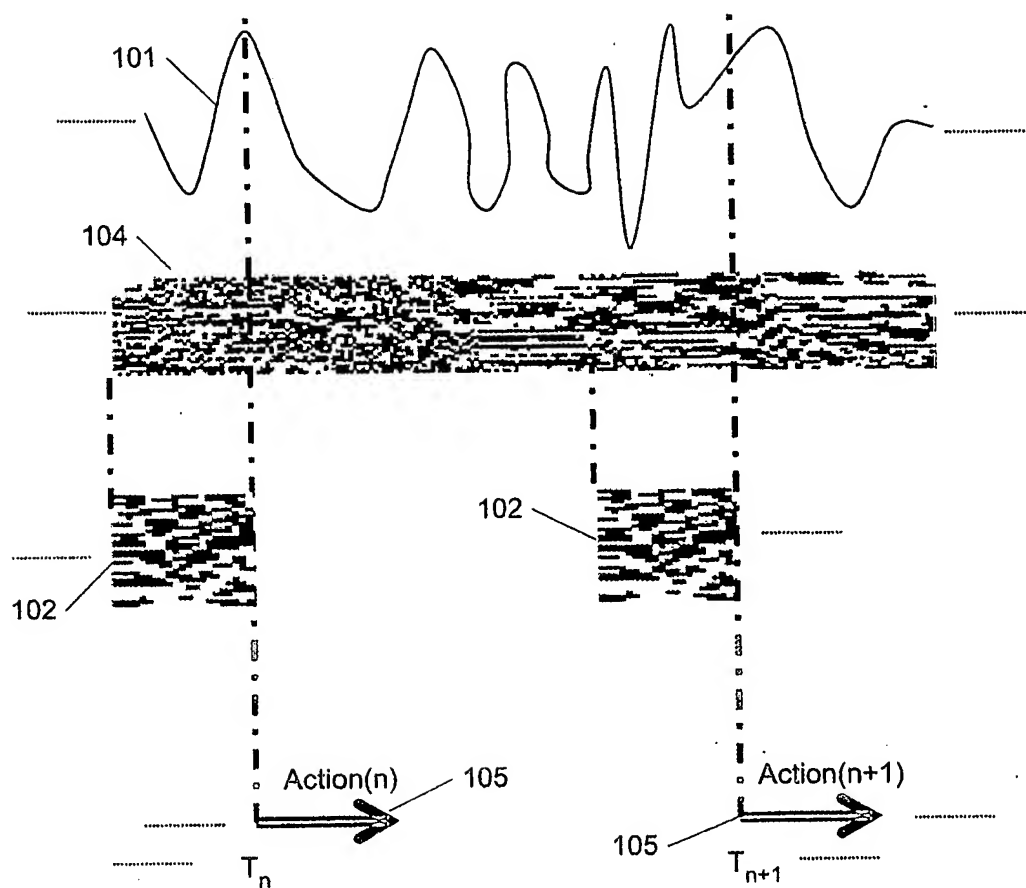
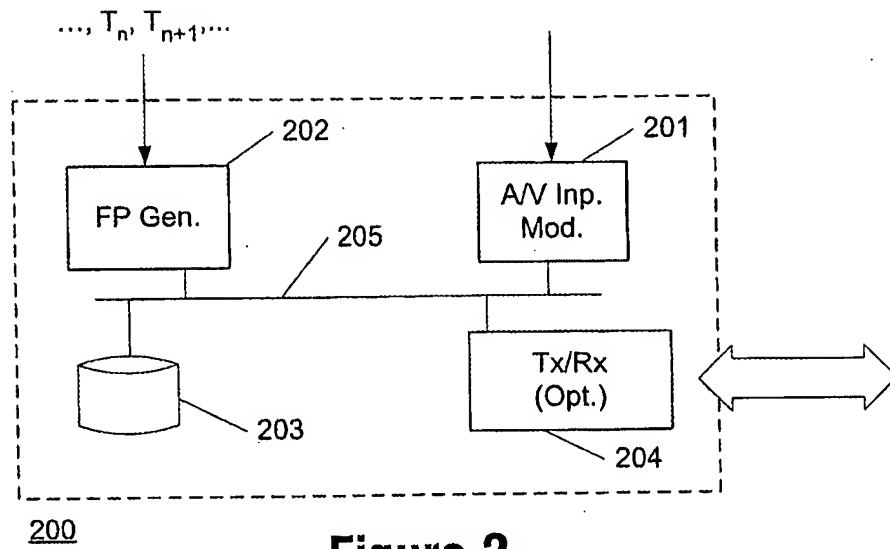
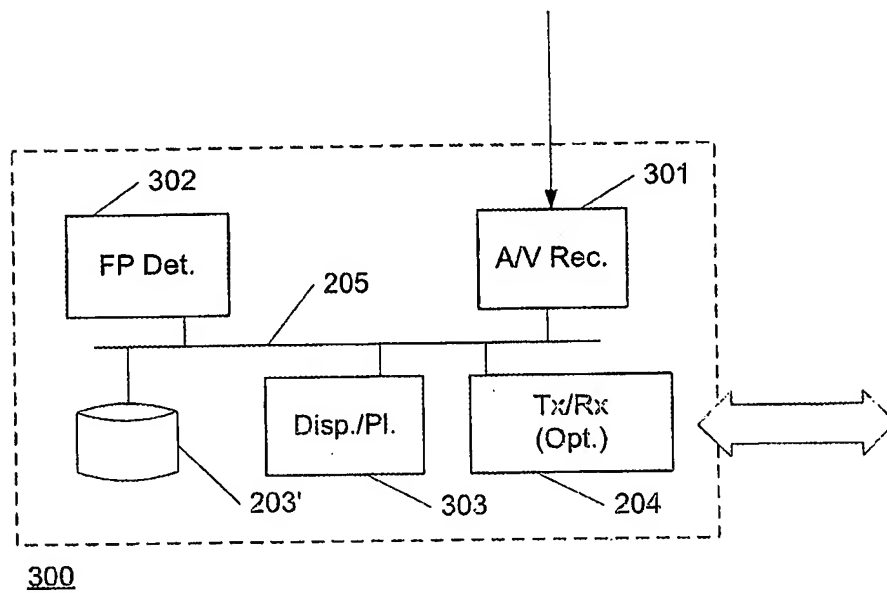


Figure 1b

3/3

**Figure 2****Figure 3**

INTERNATIONAL SEARCH REPORT

 Int. Application No
 PCT/IB2004/051128

| A. CLASSIFICATION OF SUBJECT MATTER IPC 7 H04N7/24 G06T1/00 H04N7/26 | | |
|---|--|--|
| According to International Patent Classification (IPC) or to both national classification and IPC | | |
| B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) IPC 7 H04N G06T | | |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched | | |
| Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, WPI Data, PAJ | | |
| C. DOCUMENTS CONSIDERED TO BE RELEVANT | | |
| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
| X | EP 1 054 335 A (DIGIMARC CORP) 22 November 2000 (2000-11-22) abstract page 44, paragraph 457 - page 45, paragraph 467 page 47, paragraph 473; figure 3 ----- | 1-21 |
| X | US 2002/178410 A1 (OOSTVEEN JOB CORNELIS ET AL) 28 November 2002 (2002-11-28) page 5, right-hand column, paragraph 56 - paragraph 59 ----- -/-- | 1-21 |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex. | | |
| * Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *&* document member of the same patent family | | |
| Date of the actual completion of the international search 27 October 2004 | | Date of mailing of the international search report 10/11/2004 |
| Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016 | | Authorized officer Schoeyer, M |

Form PCT/ISA/210 (second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

International Application No

PCT/IB2004/051128

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| A | KATE TEN W ET AL: "TRIGG & LINK A NEW DIMENSION IN TELEVISION PROGRAM MAKING" PROCEEDINGS OF THE EUROPEAN CONFERENCE ON MULTIMEDIA APPLICATIONS, SERVICES AND TECHNIQUES, XX, XX, May 1997 (1997-05), pages 51-65, XP000668931 cited in the application page 53, paragraph 3 - page 55, paragraph 4 | 1-21 |
| A | ----- WO 01/61987 A (WINOGRAD JOSEPH M ; PETROVIC RADE (US); VERANCE CORP (US)) 23 August 2001 (2001-08-23) abstract page 3, line 8 - page 4, line 5 | 1-21 |
| P,A | ----- WO 03/096337 A (BRUEKERS ALPHONS A M L ; KONINKL PHILIPS ELECTRONICS NV (NL); VANDER) 20 November 2003 (2003-11-20) abstract page 3, line 13 - page 5, line 28 ----- | 1-21 |

Form PCT/ISA/210 (continuation of second sheet) (January 2004)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/1B2004/051128

| Patent document cited in search report | | Publication date | Patent family member(s) | Publication date |
|---|----|---------------------|----------------------------|---------------------|
| EP 1054335 | A | 22-11-2000 | US 2003050961 A1 | 13-03-2003 |
| | | | US 6650761 B1 | 18-11-2003 |
| | | | US 6311214 B1 | 30-10-2001 |
| | | | US 2003040957 A1 | 27-02-2003 |
| | | | AU 4851300 A | 05-12-2000 |
| | | | CA 2373511 A1 | 23-11-2000 |
| | | | EP 1054335 A2 | 22-11-2000 |
| | | | EP 1185967 A1 | 13-03-2002 |
| | | | JP 2002544637 T | 24-12-2002 |
| | | | WO 0070585 A1 | 23-11-2000 |
| | | | US 2002062382 A1 | 23-05-2002 |
| | | | US 2002112165 A1 | 15-08-2002 |
| | | | US 2003105730 A1 | 05-06-2003 |
| | | | US 2003130954 A1 | 10-07-2003 |
| | | | US 6522770 B1 | 18-02-2003 |
| | | | US 2004153649 A1 | 05-08-2004 |
| | | | US 2004128514 A1 | 01-07-2004 |
| | | | US 2004190750 A1 | 30-09-2004 |
| | | | US 2001032251 A1 | 18-10-2001 |
| | | | US 2002018579 A1 | 14-02-2002 |
| | | | US 2002016816 A1 | 07-02-2002 |
| | | | US 2002131076 A1 | 19-09-2002 |
| | | | US 2002164053 A1 | 07-11-2002 |
| | | | US 2002181736 A1 | 05-12-2002 |
| | | | US 2002181737 A1 | 05-12-2002 |
| | | | US 2003012403 A1 | 16-01-2003 |
| | | | US 2002032864 A1 | 14-03-2002 |
| | | | US 2002028000 A1 | 07-03-2002 |
| | | | US 2003037075 A1 | 20-02-2003 |
| | | | US 2002170966 A1 | 21-11-2002 |
| | | | US 2003056104 A1 | 20-03-2003 |
| | | | US 2001023193 A1 | 20-09-2001 |
| | | | US 2002033844 A1 | 21-03-2002 |
| US 2002178410 | A1 | 28-11-2002 | CN 1461565 T | 10-12-2003 |
| | | | EP 1362485 A1 | 19-11-2003 |
| | | | WO 02065782 A1 | 22-08-2002 |
| | | | JP 2004519015 T | 24-06-2004 |
| WO 0161987 | A | 23-08-2001 | US 6737957 B1 | 18-05-2004 |
| | | | AU 4371601 A | 27-08-2001 |
| | | | WO 0161987 A2 | 23-08-2001 |
| | | | US 2004169581 A1 | 02-09-2004 |
| WO 03096337 | A | 20-11-2003 | WO 03096337 A2 | 20-11-2003 |

Form PCT/ISA/210 (patent family annex) (January 2004)